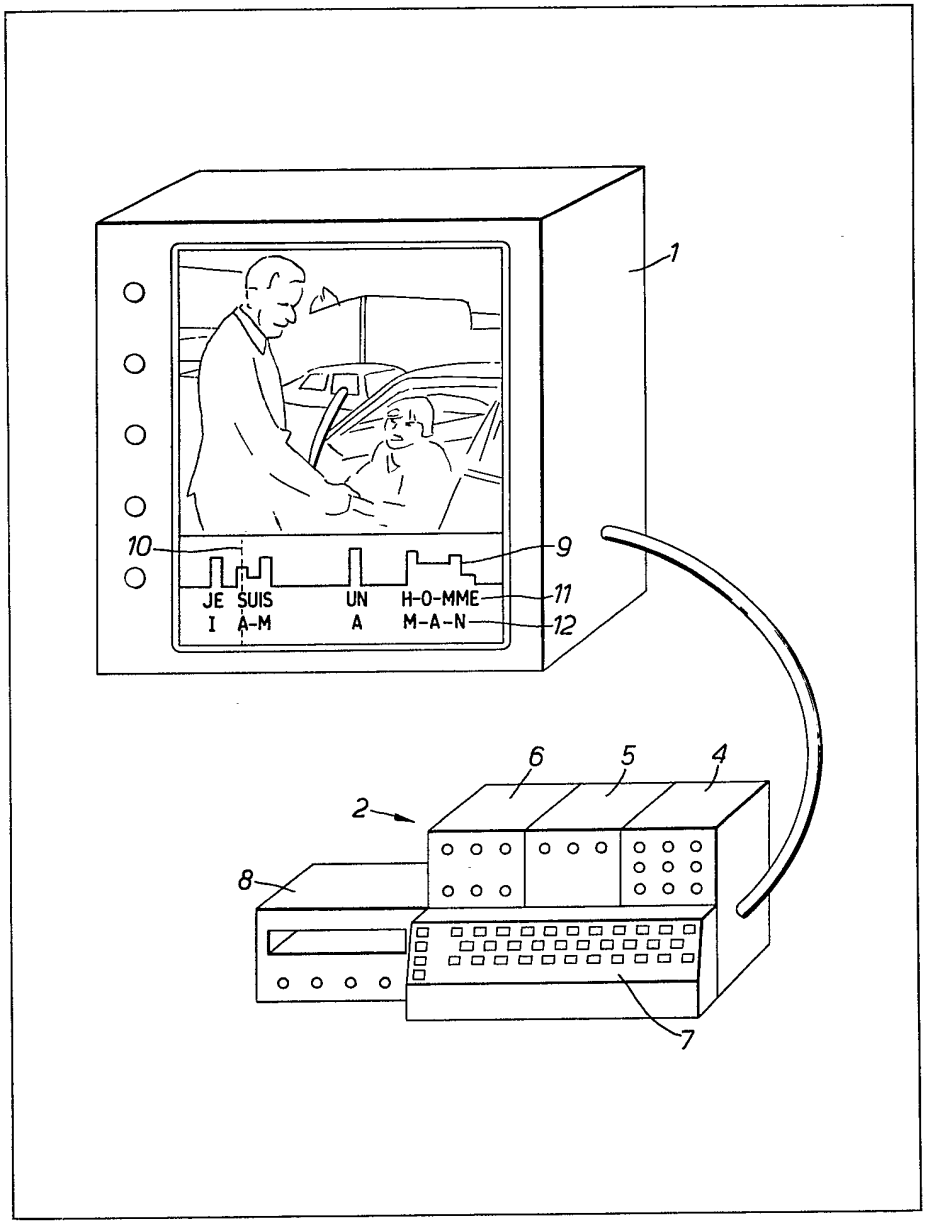
(12) UK Patent Application (19) GB (11) 2 101 795 A

(21) Application No **8219618**
(22) Date of filing **7 Jul 1982**
(30) Priority data
(31) **8120940**
(32) **7 Jul 1981**
(33) **United Kingdom (GB)**
(43) Application published **19 Jan 1983**
(51) **INT CL³ G11B 27/00 5/00**
(52) Domestic classification **G5R** B25 B264 B265 B311 B321 B36Y B373 B37Y B38Y B402 B682 B788
(56) Documents cited **None**
(58) Field of search **G5R**
(71) Applicants **John Lyndon Cross, 27 Cardiff Road, Llandaff, Cardiff, Glamorgan. Kenneth Edward Rock, 41 Allt-yr-yn Avenue, Newport, Gwent.**
(72) Inventors **Kenneth Edward Rock**
(74) Agents **Wynne-Jones Laine' and James, Morgan Arcade Chmbers, 33 St Mary Street, Cardiff, Glamorgan CF1 2AB.**

(54) **Dubbing translating of soundtracks on films**

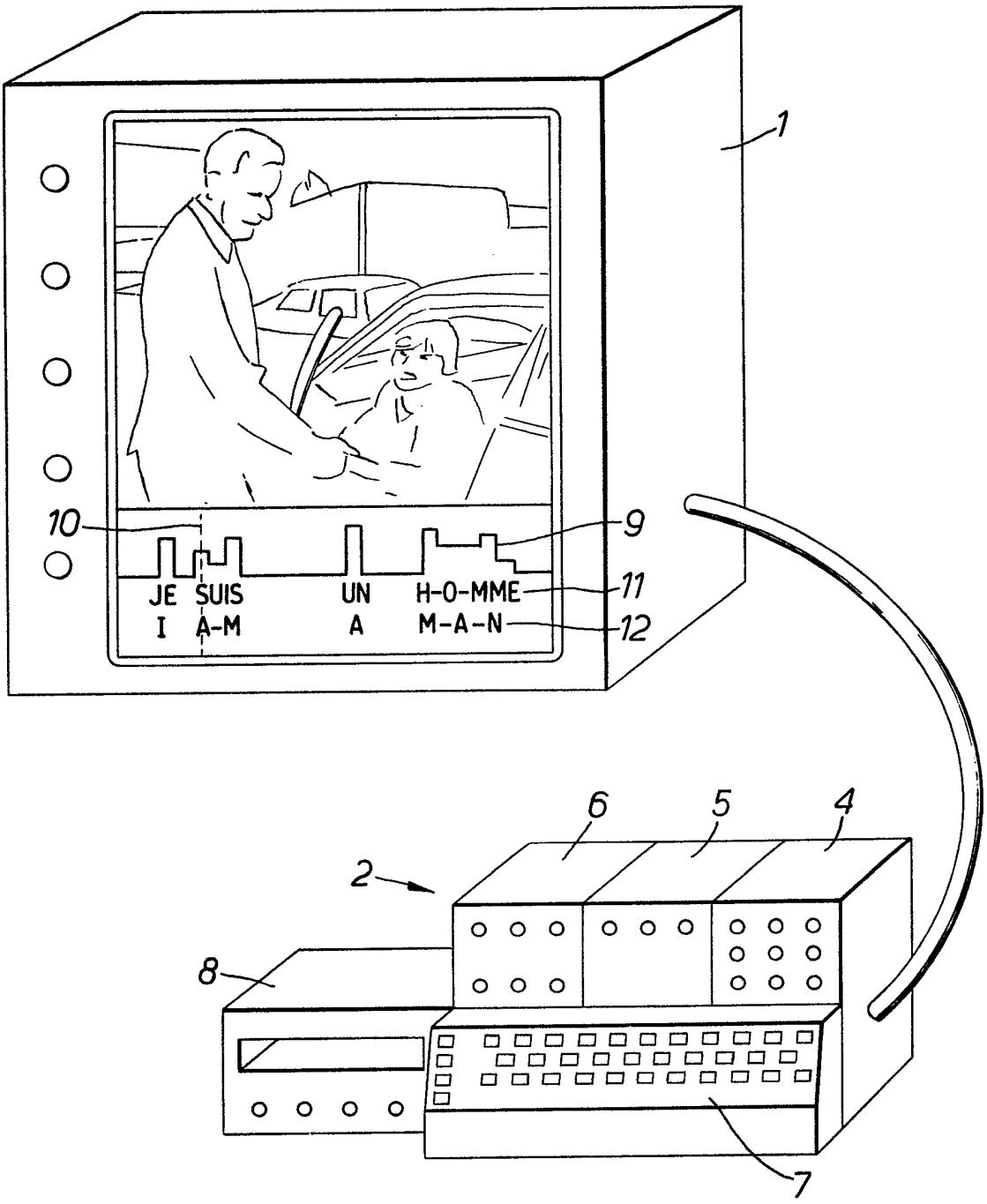
(57) A portion of a speech soundtrack 11 to be translated is displayed on a video screen of a playback unit 1 in synchronisation with lip movements of actors speaking the words with reference to a graphical sound representation histogram 9 produced by a sound conversion section 5 of a computing unit 2. A translated speech 12 is prepared and also displayed on the video screen by use of a word processor 7. When the film is played through the playback unit 1, a new actor can speak the translated speech in proper synchronisation by timing the speaking of each syllable at the time when the displayed symbol moves past a marker 10. A computation section 4 stores details relating to segments of the speech soundtrack and spacing of syllable portions is achieved by an editing section 6. New material is recorded in a recording unit 8.



GB 2 101 795 A

The drawing(s) originally filed was/were informal and the print here reproduced is taken from a later filed formal copy.

1/1



SPECIFICATION

Improvements relating to speech translation apparatus

This invention is concerned with means for producing a dubbed translation of the speech track of a film. The term "film" is used herein generally to refer to a celluloid strip film, a video tape, or other means of storing audio-visual recordings. The production of a translated speech track of course requires a number of steps, the first of which is to prepare a translated script which relates closely to the original dialogue and yet provides a sequence of words whose syllables may be matched as far as possible with the lip movements of the actors appearing on the film. The invention is more concerned with the second aspect of this procedure in providing means whereby good synchronisation of lip movement made by the original actor may be achieved by an actor speaking the translated dialogue.

A conventional method is to cut a copy of the entire film into segments which are connected as loops. Each loop is then projected on a screen and run repeatedly so that the actors can rehearse the new dialogue. When lip synchronisation is acceptable a recording is made. Other methods of aiding the actors to achieve adequate lip synchronisation have been tried in the past but these have required specialised and expensive equipment and higher technical expertise. The additional benefits did not outweigh the cost and so at present the industry still tends to use the traditional "loop" method.

It is an object of this invention to provide apparatus for enabling dubbing of a film into a new language to be carried out which is both relatively easy to use and economical of both operators' time and actors' time.

Accordingly this invention provides speech translation apparatus for the speech soundtrack of a film, comprising a video display unit having a screen for displaying the film material, a word processing unit enabling the script of the speech soundtrack or a translation thereof to be produced and displayed in alphabetical characters as a moving display on the video screen so as to pass a marker on the screen, and processing means for enabling the displayed script to be so positioned that the words pass the marker in synchronisation with the timing and lip movements of actors speaking words on the screen.

With such an apparatus all the necessary operations can be carried out from a single control console by one operator so as to produce the alphabetical display of the new speech soundtrack. It is only after this stage that the actors need to be involved and they are provided with the final product so that the film can be played and they will know precisely where syllables of words have to be spoken in order to be synchronised with lip movements of the actors on the film.

In the preferred embodiment the processing means comprises means for converting speech sounds on the soundtrack of the film into a graphical representation of sound amplitude as a moving display on the video screen so as to pass the marker on the screen in synchronisation with the soundtrack of the film. Thus the speech convertor could be designed to provide a histogram or other digital representation of the sound amplitude. Desirably the timing of the speech convertor is arranged to be controlled by a time code carried by the video tape.

As an alternative the processing means could comprise apparatus for ensuring that the words are spaced evenly within a predetermined time period. Thus for example an operator could type out the wording of a sentence (or part of a sentence), determine points related to a time code between which an actor will be speaking that sentence (or part of the sentence) and instruct the processing means to space the typed words evenly between the start and end points.

It is greatly preferred that the video display unit should include the facility of freezing the video picture. This enables words to be prepared on the screen using the word processor whilst the picture is not moving. It is also preferred that the word processor and/or the video unit should be provided with one or more controls for freezing the film display, shifting alphabetical characters on the screen and varying the spacing of alphabetical characters on the screen.

The apparatus will desirably include a computing unit which is able to control electronically the division of the video recording into segments and to play back any of the segments at will and store information relating to each segment. Advantageously the computer unit will also be able to store details of any graphical representation of the sound amplitude, positioning of the alphabetical character script produced by the word processor and general information relating to each segment.

In order to reduce the amount of storage space within the apparatus (such as in the computing unit when present) it is advantageous to include a storage recorder for providing a permanent store on a replayable medium of the positioning of the alphabetical character script relative to a time code. The apparatus will also desirably include a recording unit for producing a permanent replayable record of the alphabetical characters for a newly recorded speech in a new language produced by the word processor relative to the time code.

The invention also extends to a method of translating a soundtrack for a film using speech translation apparatus as hereinbefore defined, wherein a video tape is prepared carrying video material, the soundtrack and a time code, the script of the speech soundtrack is displayed on the screen by using the word processor and processing means to position alphabetical characters in synchronisation with the timing and lip movements of actors speaking words on the screen, a suitable translation of the speech is prepared and similarly displayed on the screen using the word processor and the translated speech is recorded vocally so as to be synchronised with the script reproduced on the video screen as the video tape is played through the display unit.

Preferably a graphical representation of sound amplitude of the soundtrack is created as a moving display on the video screen to pass a marker in

synchronisation with the soundtrack of the film. It is possible for the speech to be recorded first and then be electronically shaped to fit lip movements and/or the sound amplitude graphical representation by a technician, so that the actor can produce a more natural speech.

The invention may be performed in various ways and one preferred embodiment thereof will now be described with reference to the accompanying drawing which illustrates equipment which may be used for preparing a translation of the soundtrack for a film.

The apparatus illustrated in the drawings generally comprises a video playback machine 1 and a computing unit 2. The playback unit 1 is associated with a recorder for playing a video cassette. The material on the tape of the cassette will have been recorded from original film material so as to carry the original picture material on one track, the speech (but excluding other sound effects) on a second track and a time code on a third track which provides suitable access times to the video material and soundtrack. If necessary the time code could be produced by a series of coded dots on the video track. Subsequent operations in preparing a translation and recording the actors' voices speaking the translated material is carried out in relation to the video tape so that the original material is spared further degradation and the system operation is speeded up since the equipment (as shown in the drawings) is standardised.

The computing unit 2 comprises five basic section, namely a computation section 4, a sound conversion section 5, an editing section 6, a word processor 7 and a recording unit 8. The various parts of the computing unit operate in relation to the time code which is carried by the tape on the video cassette.

The first task is to break down the original text into usable segments. The video playback unit 1 incorporates controls whereby the picture may be frozen and segments of the track may be repeated continuously as a "loop". The operator (who may also be the translator) selects each segment by inspecting the displayed picture and presses buttons controlling the computation section to identify relevant "in" and "out" times (whilst running the recorder at slow speeds if necessary). A record will also be made of the actors necessary to revoice each segment. At the end of this process the track will be divided into a sequence of numbered segments each covering from ten to thirty seconds of speech or effects, the numbering of the segments and details of actors appearing in each segment being stored in the computation section 4.

Now that the material can be referred to in the form of segment numbers progress can be made in preparing a translation of the spoken material occurring in each segment. The segments can be chosen in any preferred order. For example if a particular actor has a limited availability all the segments for which that actor will be required can be dealt with first. For each segment being analysed the following procedure will be carried out. Firstly the required segment is selected by typing in the segment number (on the word processor consul 7) and the computation section will then cause that segment to be displayed on the playback unit 1 as a repeating loop. For ease of use the segment will be played back with a run-up time of five seconds and there will be a built-in cue to assist the operator. As the segment is played forward the speech track is analysed by feeding segments to shaping circuits and an analogue to digital convertor in the sound conversion section 5 so as to create a sequence of eight-bit numbers related to certain predetermined threshold levels and a graphic character is stored in section 5 dependent upon the amplitude of the incoming wave form and the sequence of graphic characters is displayed on the playback unit in the form of a histogram. The histogram is keyed in time code and will move across the segment of the playback unit so that the part of the histogram passing a marker on the screen will be synchronised with the speech syllable which is being spoken by the actor on the film at that precise moment.

It should be noted here that, generally speaking, normal speech is delivered at about five syllables per second. This speed is limited by intelligibility and is the same for any language. Another factor is that the video sampling rate (equivalent to film frame rate) is twenty five per second so that for an average syllable lasting one fifth of a second there are five "frames" associated with that syllable. Thus the storage of the graphic characters for the histogram of a segment lasting thirty seconds will require thirty times twenty five bytes of memory (i.e. seven hundred and fifty bytes in total). Thus a record, accurate to a frame, is stored which describes the relationship of the syllables uttered to time code on the screen.

Once the graphic characters are stored in the sound conversion section 5, they can be produced onto the screen of the playback unit 1 and made to move to the left as a histogram 9 towards the marker 10 on the screen. The translator will have a copy of the spoken text and will be able to introduce this text onto the screen beneath the train of graphic characters using the word processor keyboard 7. Generally this will be done by freezing the picture and introducing the syllables of the text, in the form of alphabetic characters, directly beneath the graphic characters depicting the same lip movement in the original language, as illustrated on the line 11. Once the letters have been typed in the editing section 6 can be used to modify the spacing and positioning of the letters so that they lie directly below the graphic characters of the histogram 9. Once the original text is located the same procedure can be carried out to introduce a translated text as shown for example on the line 12. The word processor 7 can be used, in association with the editing section 6, to change the words or spacing of the translated text as desired. When a suitable translation has been correctly positioned in relation to the histogram 9, the translated text and the positioning of the alphabetic characters of that text will be stored in the computing unit in relation to the time code.

When the translated text is located in the optimum position, the stored information for a particular segment can be recorded on a permanent storage

medium such as a "floppy disc" so that the memory does not become overloaded. For each segment which has been dealt with in this way, the video playback unit can be operated, in synchronism with the material carried by the floppy disc to display the words of the translated text in the predetermined spacing (with or without the histogram 9) as the picture is displayed. The words will pass across the screen from right to left and the actor will know that he has to speak each word or syllable as the relevant alphabetic characters pass the marker 10. There is a two second delay as the material passes across the screen which enables the actor to predict the arrival of the words against the marker 10 so that he can achieve lip synchronisation as the travelling words reach this cue point. When it is felt that lip synchronisation is acceptable the segment of the newly recorded speech in the translated language can be permanently recorded on a new storage medium and the next segment can be dealt with.

A modified method of obtaining a new soundtrack with the translated speech is to pass the script to the actors who will read the translated words to achieve the best effect. Timing is not critical and this allows the actors to concentrate on the mood of the play. The actor's speech is then digitised via an analogue to digital convertor and the words can then be expanded or compressed to fit the original voice prints as illustrated by the histogram 9. The modified digitally encoded voice is then played back via a digital to analogue convertor and is synchronised by means of the time code to be mixed down for the final tape.

This system provides a number of useful features as well as those apparent from the previous description. Firstly there is no need to interfere with the original film since the film itself does not need to be cut into segmental strips; so the quality of the film is maintained. Instead the computation section 4 electronically determines each segment of ten to thirty seconds duration. The computer can be used to record information relative to each segment, namely the start and end points, the code numbers for each actor and a tabulation of the respective segments. This allows groups of segments to be chosen so that a particular actor can be selected by his code and provided with a list of the segments assigned to him. Optimum use can then be made of the varying availability of the actors. The use of a floppy disc or the like for permanent storage of the prepared material for displaying the new language for each segment avoids overloading of the computer storage. The number of alphabetic characters required to determine a particular syllable will of course vary and some languages employ a more economical use of characters than others. However as a general rule an average syllable will require four alphabetic characters and since the time span for each average syllable is about one fifth of a second there are five frames within which the four alphabetic characters may be placed which is more than enough space.

As an alternative to displaying the histogram 9 the apparatus could be used by a skilled operator to position the original and translated script in such a way that the words will be synchronised with lip movements of the actors displayed on the screen. In this case the operator would project the film material through slowly, frame by frame, until an actor starts to speak a particular phrase or sentence whereupon the operator will press a button to indicate the start point to the control equipment. A separate button will be pressed to indicate the end point (for the sentence or phrase being spoken) relative to a time code, thus determining the temporal positiion of that particular part of the spoken script. The control equipment will incorporate processing means which will cause the typed words of the script to be evenly spaced, as they are displayed on the screen, over a predetermined time period. Once the script has been positioned in this way the translated script can be typed so as to situate below the original script in the required manner and the subsequent stages of dubbing the film will continue as before.

CLAIMS

1. Speech translation apparatus for the spech soundtrack of a film, comprising a video display unit having a screen for displaying the film material, a word processing unit enabling the script of the speech soundtrack or a translation thereof to be produced and displayed in alphabetical characters as a moving display on the video screen so as to pass a marker on the screen, and processing means for enabling the displayed script to be so positioned that the words pass the marker in synchronisation with the timing the lip movements of actors speaking words on the screen.

2. Speech translation apparatus according to claim 1, wherein the processing means comprises means for converting speech sounds on the sound track of the film into a graphical representation of sound amplitude as a moving display on the video screen so as to pass the marker in synchronisation with the soundtrack of the film.

3. Speech translation apparatus according to claim 2, wherein the speech convertor is designed to provide a histogram or other digital representation of the sound amplitude.

4. Speech translation apparatus according to claim 2 or claim 3, wherein the timing of the speech convertor is arranged to be controlled by a time code carried by the video tape.

5. Speech translation apparatus according to claim 1, wherein the processing means comprises apparatus for ensuring that the words are spaced evenly within a predetermined time period.

6. Speech translation apparatus according to any one of claims 1 to 5, wherein the video display unit includes the facility of freezing the video picture.

7. Speech translation apparatus according to any one of claims 1 to 6, wherein the word processor and/or the video unit is provided with one or more controls for freezing the film display, shifting alphabetical characters on the screen and varying the spacing of alphabetical characters on the screen.

8. Speech translation apparatus according to any one of claims 1 to 7, including a computing unit which is able to control electronically the diversion of the video recording into segments and to play

back any of the segments at will and store information relating to each segment.

9. Speech translation apparatus according to claim 8, wherein the computer also stores details of any graphical representation of the sound amplitude, positioning of the alphabetical character script produced by the word processor and general information relating to each segment.

10. Speech translation apparatus according to any one of claims 1 to 9, including a storage recorder for providing a permanent store or a replayable medium of the positioning of the alphabetical character script relative to a time code.

11. Speech translation apparatus according to any one of claims 1 to 10, including a recording unit for producing a permanent replayable record of the alphabetical characters for a newly recorded speech in a new language produced by the word processor relative to the time code.

12. Speech translation apparatus substantially as herein described with reference to the accompanying drawings.

13. A method of preparing a translated soundtrack for a film using speech translation apparatus as claimed in any of claims 1 to 12, wherein a video tape is prepared carrying video material, the soundtrack and a time code, the script of the speech soundtrack is displayed on the screen by using the word processor and processing means to position alphabetical characters in synchronisation with the timing and lip movements of actors speaking words on the screen, a suitable translation of the speech is prepared and similarly displayed on the screen using the word processor, and the translated speech is recorded vocally so as to be synchronised with the script reproduced on the video screen as the video tape is played through the display unit.

14. A method according to claim 13, wherein a graphical representation of sound amplitude of the soundtrack is created as a moving display on the video screen to pass a marker in synchronisation with the soundtrack of the film.

15. A method according to claim 13 or claim 14, wherein the speech is first recorded vocally and is then electronically shaped to fit lip and/or the sound amplitude graphical representation by a technician.

16. A method of preparing a translated soundtrack for a film substantially as herein described with reference to the accompanying drawings.

Printed for Her Majesty's Stationery Office, by Croydon Printing Company Limited, Croydon, Surrey, 1983.
Published by The Patent Office, 25 Southampton Buildings, London, WC2A 1AY, from which copies may be obtained.